



XML

基础教程与实验指导

► 66个典型范例，与技术讲解有机结合，
105道上机实验题，巩固知识，便于演练



XML

基础教程与实验指导

宋武 编著

清华大学出版社
北 京

内 容 简 介

可扩展标记语言 XML 是网络应用开发的一项技术,它同 HTML 一样,是一种标记语言。本书是学习和应用 XML 语言的实用教材,书中详细阐述了 XML 的基本概念、语法规则、文档类型定义、模式定义、级联样式表、可扩展样式表、文档对象模型,此外还介绍了 XML 在 Java 和 .NET 中的编程,最后通过一个综合案例演示了在实际项目开发中 XML 的应用。本书内容由浅入深,在讲解基本概念和基础知识的同时给出了大量实例,每章还给出了填空题、选择题和上机题,便于读者巩固所学的知识。

本书可作为高等院校计算机、电子商务以及信息类相关专业的教材,也可供相关技术人员、广大 Web 应用程序开发者和用户参考。

本书封面贴有清华大学出版社防伪标签,无标签者不得销售。

版权所有,侵权必究。侵权举报电话:010-62782989 13701121933

图书在版编目(CIP)数据

XML 基础教程与实验指导/宋武编著. —北京:清华大学出版社,2013

ISBN 978-7-302-32967-1

I. ①X… II. ①宋… III. ①可扩展语言—程序设计 IV. ①TP312

中国版本图书馆 CIP 数据核字(2013)第 147784 号

责任编辑:夏非彼

封面设计:王 翔

责任校对:闫秀华

责任印制:沈 露

出版发行:清华大学出版社

网 址: <http://www.tup.com.cn>, <http://www.wqbook.com>

地 址:北京清华大学学研大厦 A 座 邮 编:100084

社总机:010-62770175 邮 购:010-62786544

投稿与读者服务:010-62776969, c-service@tup.tsinghua.edu.cn

质 量 反 馈:010-62772015, zhiliang@tup.tsinghua.edu.cn

印 装 者:三河市金元印装有限公司

经 销:全国新华书店

开 本:190mm×260mm

印 张:22.25

字 数:585 千字

版 次:2013 年 8 月第 1 版

印 次:2013 年 8 月第 1 次印刷

印 数:1~4000

定 价:42.00 元

产品编号:050341-01

前 言

XML 是 eXtensible Markup Language 的简写，它是一种扩展性标识语言。XML 在 Web 编程、新型数据库系统、计算机网络应用编程、网络数据交换和跨平台编程中发挥着越来越重要的作用。XML 正在成为电子商务运营和数据管理的核心技术。

本书系统介绍了设计和使用 XML 文档所必需的基本技术，以及很多现在非常重要的相关技术。通过本书，可使读者对 XML 语言有一个全面的了解，掌握相关的主要知识点，在软件设计中运用 XML 技术灵活地进行信息的描述、交换和显示。全书共分为 12 章，具体内容如下。

第 1 章 XML 概述

本章主要介绍了标记语言产生的原因和发展、SGML 语言、HTML 超文本语言，最后介绍了 XML 语言的概念、优势和相关技术。

第 2 章 XML 语法

本章主要介绍 XML 的语法，重点介绍了 XML 文档结构，XML 格式的约束规则，如何声明元素和属性以及怎样使用名称空间等。

第 3 章 XML 开发工具

本章主要介绍在整个 XML 技术的发展阶段出现过的编辑器、解析器和浏览器及其他工具软件，并对 XMLSpy2011 的安装、主界面、主要功能和基本的使用进行了详情的介绍。

第 4 章 XML 文档类型定义

本章主要介绍的是确保 XML 文档有效的文档类型定义，重点介绍了 DTD 中元素、属性、实体的声明语法和格式，以及在 XML 文档中使用 DTD 的方式。

第 5 章 XML 模式

本章主要介绍了 XML 模式定义的基本语法知识，重点介绍了模式的文档结构，XML 模式中元素及属性的声明语法和使用 XML 模式的方式。

第 6 章 用 CSS 显示 XML

本章首先介绍了 CSS 的作用，然后是 CSS 语法及常见的 CSS 属性的设置，并以实例的方式讲述了在 XML 中如何使用 CSS。

第 7 章 XSL 转换

本章讲述了 XSL 的基本概念，着重讲解 XSL 转换语言，XSL 文档的编写，模板的规则及应用模板，XSL 的转换执行过程，实现 XSL 的方式。

第 8 章 数据岛

本章主要介绍了数据岛技术，详细介绍了数据岛的使用方式、数据源对象 DSO 的操作，并给出了遍历和分页显示 XML 文档的实例

第 9 章 XML 链接

本章主要介绍构成 XML 链接语言的两大部分——XLink 和 XPointer 语言的相关概念和具体使用方法。

第 10 章 DOM 和 SAX

本章重点介绍文档对象模型 DOM 的基本概念、常用接口、常用对象和 DOM 与 XML 如何进行交互。此外，还讲述了操作 XML 文档的另外一种方式 SAX，包括其工作原理、常用的事件和接口以及它对 XML 的解析。

第 11 章 .NET 中的 XML 编程

本章主要介绍了在 .NET 平台上操作和处理 XML 文件要用到的具体技术及相关对象，重点介绍了 .NET 平台上的 DOM 编程方法，以及 XML 与 ADO.NET 技术的关系。

第 12 章 基于 XML 的新闻发布系统

在本章中，通过一个综合案例演示了在实际项目开发中 XML 的具体应用，通过对案例的剖析，体会 XML 技术在 Web 应用方面的优势并进一步掌握利用 DOM 操作 XML 文档的方法和技巧。

本书技术阐述与实践应用相结合，强调理论结合实际需要，始终以介绍 XML 中已成熟的标准和应用技术为主。书中的应用实例均来自于实际开发中的需要，读者可以对其稍加修改后直接应用到实际开发中。

本书可作为高等院校计算机、电子商务以及信息类相关专业的教材，也可供相关技术研究人员、广大 Web 应用程序开发者和用户参考。

本书主要由宋武主持编写，此外，陈小亮、张国栋、张国华、李华、王林、李志国、陈晨、冯慧、徐红、吴文林、周建国、张建、刘海涛、张琴、高梅、吴晓、朱维、陈浩、汪梅、姚琳、何武和许小荣等同志也参加了本书的编写，在整理材料方面给予了编者很大的帮助，在此，编者对他们表示衷心的感谢。

由于作者水平有限，书中错误、纰漏之处难免，欢迎广大读者、同仁批评斧正。

作 者
2012 年 9 月

目 录

第 1 章 XML 概述	1
1.1 标记语言	1
1.1.1 标准通用标记语言 (SGML)	1
1.1.2 超文本标记语言 HTML	2
1.1.3 可扩展标记语言 (XML)	3
1.2 HTML 概述	5
1.2.1 HTML 基本结构	5
1.2.2 HTML 常用标记	6
1.3 XML 简介	9
1.3.1 XML 和 SGML、HTML 的关系	9
1.3.2 XML 的简单例子	10
1.3.3 XML 的优势	12
1.3.4 XML 的应用	14
1.3.5 XML 的缺陷	17
1.4 XML 相关技术	17
1.4.1 文档类型定义 (DTD)	18
1.4.2 层叠样式表 (CSS)	18
1.4.3 可扩展样式语言 (XSL)	19
1.4.4 文档对象模型 (DOM)	19
1.4.5 XML Schema	20
1.4.6 XLink 与 XPointer	20
1.5 习题	20
第 2 章 XML 语法	23
2.1 XML 文档的结构	23
2.2 XML 文档的序言	25
2.2.1 XML 声明	25
2.2.2 XML 处理指令	26
2.2.3 XML 注释	27
2.3 XML 文档元素	28
2.3.1 元素	28
2.3.2 XML 元素的属性	31
2.4 CDATA	33
2.4.1 特殊字符的使用	33
2.4.2 CDATA 的格式	34
2.4.3 CDATA 的位置	35
2.5 XML 命名空间	36

2.5.1	命名空间的概念	36
2.5.2	定义命名空间	37
2.5.3	命名空间的使用	38
2.5.4	默认命名空间	39
2.6	空白字符	40
2.7	XML 文档的数据结构	42
2.8	正确的 XML 文档	43
2.8.1	格式良好的 XML 文档	44
2.8.2	有效的 XML 文档	44
2.9	习题	45
2.10	上机实验	46
第 3 章	XML 开发应用环境	50
3.1	XML 软件	50
3.1.1	XML 浏览器	50
3.1.2	XML 解析器	52
3.1.3	XML 编辑器	54
3.2	XMLSpy	56
3.2.1	XMLSpy 的安装	56
3.2.2	XMLSpy 的主界面	58
3.2.3	XMLSpy 的主要功能	62
3.2.4	XMLSpy 的使用	66
3.3	习题	67
3.4	上机实验	68
第 4 章	XML 文档类型定义	70
4.1	DTD 的作用	70
4.2	DTD 的结构	71
4.3	DTD 中的元素	73
4.3.1	元素的声明	73
4.3.2	元素的基本类型	74
4.3.3	元素限定符	77
4.4	DTD 属性	79
4.4.1	属性的声明	80
4.4.2	属性的类型	81
4.4.3	属性的附加声明	90
4.5	DTD 的引用	92
4.5.1	内部 DTD	92
4.5.2	外部 DTD	92
4.6	DTD 中的实体	97
4.6.1	实体的概念和分类	97
4.6.2	普通实体	97

4.6.3	参数实体	100
4.6.4	不可解析实体	102
4.7	习题	103
4.8	上机实验	104
第 5 章	XML 模式	108
5.1	Schema 概述	108
5.2	XML Schema 的基本格式	109
5.3	XML Schema 元素的声明	113
5.3.1	XML Schema 根元素	113
5.3.2	element 元素	114
5.3.3	空元素的定义	115
5.3.4	设置默认值和固定值	116
5.3.5	元素的引用	117
5.3.6	元素的替换	118
5.4	XML Schema 的数据类型	119
5.4.1	简单数据类型	119
5.4.2	复杂数据类型	123
5.5	XML Schema 命名空间	129
5.5.1	XML Schema 命名空间类型	129
5.5.2	引用 XML Schema 命名空间	130
5.6	XML Schema 文档的可视化创建	131
5.7	习题	134
5.8	上机实验	135
第 6 章	用 CSS 显示 XML	139
6.1	CSS 概述	139
6.1.1	CSS 显示 XML 基础	139
6.1.2	CSS 显示 XML 的实例	140
6.2	CSS 显示 XML 的基本语法	142
6.2.1	CSS 选择器	142
6.2.2	CSS 属性设置	147
6.3	CSS 链接到 XML	162
6.3.1	常规的链接指令	162
6.3.2	@import 指令	163
6.3.3	无需链接的内部 CSS	164
6.4	CSS 中的注释	165
6.5	习题	165
6.6	上机实验	166
第 7 章	XSL 转换	169
7.1	XSL 概述	169

7.1.1	XSL 的作用	169
7.1.2	XSL 的组成	170
7.1.3	XSL 的转换原理	171
7.2	XSL 文档的结构	171
7.3	XSL 模板	174
7.3.1	XSL 模板定义	174
7.3.2	其他常用属性	176
7.3.3	XSL 节点的选择	177
7.4	XSL 控制指令	182
7.4.1	简单判断指令	182
7.4.2	多条件判断指令	182
7.4.3	循环处理指令	185
7.4.4	输出内容排序	186
7.5	XPath 简介	188
7.5.1	XPath 中的节点类型和关系	188
7.5.2	XPath 的类型和运算符	190
7.5.3	XPath 的位置路径表达式	191
7.5.4	XPath 的标准函数	194
7.6	习题	196
7.7	上机实验	197
第 8 章	数据岛	201
8.1	数据岛简介	201
8.1.1	数据岛的含义	201
8.1.2	数据岛绑定	202
8.1.3	数据岛绑定的优势	203
8.1.4	支持数据绑定的 HTML 元素	204
8.2	数据岛的形式	205
8.2.1	内部数据岛	205
8.2.2	外部数据岛	206
8.3	数据岛的处理	206
8.3.1	表格对象的绑定	207
8.3.2	实现分页表格	209
8.3.3	访问 XML 元素属性	211
8.4	DSO 数据源对象	212
8.4.1	DSO 简介	212
8.4.2	访问和操作 DSO 记录	215
8.4.3	使用 DSO 显示图片	219
8.4.4	DSO 和 XSL 切换样式	221
8.5	习题	223
8.6	上机实验	224

第 9 章 XML 链接	228
9.1 什么是 XML 链接	228
9.2 XLink 语言	229
9.2.1 XLink 中的专用名词	229
9.2.2 简单链接	230
9.2.3 扩展链接	234
9.2.4 外联链接	242
9.2.5 链接库	243
9.3 XPointer 语言	245
9.3.1 XPointer 概述	245
9.3.2 XPointer 的 3 种形式	246
9.3.3 位置路径和位置集合	248
9.3.4 XPointer 对 XPath 的扩展	249
9.4 习题	252
9.5 上机实验	253
第 10 章 XML 文档对象模型	256
10.1 XML 文档解析	256
10.1.1 XML 解析器	256
10.1.2 常用解析器接口	257
10.2 DOM 概述	258
10.2.1 DOM 的结构	259
10.2.2 DOM 对象	261
10.3 DOM 的使用	264
10.3.1 JAXP 简介	264
10.3.2 JAXP 提供的类和方法	265
10.3.3 遍历 XML 文档	267
10.3.4 访问属性节点	269
10.3.5 动态创建 XML 文档	271
10.3.6 文本节点的操作	272
10.3.7 删除节点	274
10.4 SAX 概述	276
10.4.1 SAX 接口	276
10.4.2 SAX 工作机制	278
10.4.3 SAX 常用接口	280
10.5 SAX 解析器的使用	281
10.5.1 SAX 的创建步骤	281
10.5.2 SAX 解析 XML 文档	282
10.6 习题	285
10.7 上机实验	285

第 11 章 .NET 和 XML	289
11.1 .NET Framework 介绍	289
11.1.1 .NET 构架	289
11.1.2 从 MSXML 到 .NET 的 XML	290
11.1.3 XML 分析模式	290
11.2 使用流模式处理 XML	291
11.2.1 和 XML 相关的类	291
11.2.2 读取 XML	293
11.2.3 创建 XML	296
11.3 使用 DOM 处理 XML	298
11.3.1 XML 文档的保存	299
11.3.2 将 XML 读入文档	300
11.3.3 选择节点	301
11.3.4 新节点的创建	302
11.3.5 XML 文档的修改	305
11.3.6 XML 文档的删除	307
11.4 ADO.NET 和 XML	308
11.4.1 XML 与 DataSet 的关系	309
11.4.2 DataSet 访问 XML	310
11.5 习题	313
11.6 上机实验	314
第 12 章 基于 XML 的新闻发布系统	317
12.1 系统分析与设计	317
12.1.1 系统需求分析	317
12.1.2 系统模块设计	318
12.2 系统 XML 文件设计	318
12.2.1 User.xml	319
12.2.2 XMLFile.xml	319
12.3 前台管理模块的设计	320
12.3.1 前台首页的显示	321
12.3.2 新闻类别信息列表显示	323
12.3.3 新闻详情显示	326
12.3.4 查询新闻	327
12.4 后台管理模块的设计	329
12.4.1 管理新闻	329
12.4.2 添加新闻	332
12.4.3 用户管理	333
12.5 登录模块的设计	338
12.6 上机实验	339
习题答案	342

第 1 章

XML 概述

XML 的英文全称是 eXtensible Markup Language，中文翻译为“可扩展标记语言”。它是网络应用开发的一项新技术。XML 同 HTML 一样，是一种标记语言，但是 XML 的数据描述的能力比 HTML 强很多，XML 能够描述所有已知和未知的数据。XML 扩展性比较好，可以为新的数据类型制定新的数据描述规则，作为对标记集的扩展。XML 出现以后就迅速走红，目前已经成为在不同系统之间进行数据交换的基础。XML 的商用前景之所以非常广阔，也是因为它满足了当前商务数据交换的需求。

本章主要讲解关于标记语言的相关基础知识及其发展历史，同时，对目前的主流标记语言 HTML 和 XML 做了一个总括式的介绍，特别是 XML 的优缺点、应用领域和与 XML 密切相关的技术，让读者有一个大概的了解，从而为以后的学习打好基础。

1.1 标记语言

标记语言，又称为“置标语言”，是一种将文本（Text）以及与文本相关的其他信息结合起来，展现出关于文档结构和数据处理细节的电脑文字编码。与文本相关的其他信息包括文本的结构和表示信息等，它们与原来的文本结合在一起，使用标记（markup）进行标识，标记最早用于出版业，是作者、编辑以及出版商之间用于描述出版作品排版格式的。

目前，标记语言被广泛应用于网页和网络应用程序中。使用最多的标记语言是超文本标记语言（HyperText Markup Language, HTML）和可扩展标记语言（eXtensible Markup Language, XML）。

1.1.1 标准通用标记语言（SGML）

20 世纪 60 年代，为了促进数据交换和操作，IBM 公司的研究人员创建了 GML（Generalized Markup Language，通用标记语言），用来解决不同计算机系统中文档格式不相同的问题。GML 是一种可以创建各种数据标记的语言，能够方便地标识出任何数据集合的结构，因而成为许多 IBM 电子文档系统的基础。

经过多年的发展，1986 年 ISO（国际标准化组织）批准 GML 为国际标准，并正式定名为 SGML。SGML（Standard Generalized Markup Language，通用标识语言标准）是国际上定义电子文件结构和内容描述的标准。它是用于定义其他语言的元语言，以逻辑化和结构化的方式来描述文本化的文档，主要用于电子文档的创建、存储及分发。一个 SGML 文档由语法定义、文档类型定义和文档实例 3 个部分组成。语法定义部分为文档类型和文档实例定义了语法结构；文档类型定义部分定义了文档实例的逻辑结构和元素类型；文档实例部分则包含了该文档

的所有数据元素，是 SGML 文档的主体部分。

正如其名称所言，SGML 是一种采用标记来描述文档数据的通用语言，SGML 定义了基本的语法，但也允许用户建立自己的元素标记。要利用 SGML 来描述特定文档，用户必须创造一组适当的元素标记。例如，要描述有关书籍信息的文档，就可能需要命名并使用名为 BOOK、PART、CHAPTER、INTRODUCTION、SECTION 的元素标记。

SGML 是与平台无关的、结构化的、可扩展的标记语言。SGML 文档包含了一系列的 DTD (Document Type Definition, 文档类型定义) 内容，而在 DTD 中则定义了各种标记的含义，因而 SGML 的语法是可以扩展的。

SGML 文档虽然已经在电信、航空公司、汽车公司、化工企业和军队使用多年，然而，W3C (World Wide Web Consortium, 万维网联盟) 中的成员却认为，SGML 过于复杂和麻烦，难以学习和掌握，无法实现在网络上有效率传递信息。SGML 的过多弹性与其他一些特性，使编写负责在浏览器中处理与显示 SGML 信息的软件变得非常困难。鉴于这些原因，Web 发明者欧洲粒子物理实验室的研究人员根据当时的情况，在 SGML 的基础上提出了超文本标记语言 HTML。

1.1.2 超文本标记语言 HTML

HTML (Hyper Text Markup Language, 超文本标记语言)，是目前网络上应用最为广泛的语言，也是构成网页文档的主要语言。

1989 年，欧洲粒子物理实验室的研究人员开发了一个基于 SGML 的超文本简化版本，被称为超文本标记语言 (HTML)。HTML 是 SGML 的一个子集，继承了 SGML 的许多重要特点。HTML 只使用了 SGML 中的很小一部分标记，并且为了便于在计算机上实现，HTML 可使用的标记是固定的，也不需要包含 DTD，因而其语法是不可扩展的。HTML 这种固定的语法易学易用，开发用来处理 HTML 文档的浏览器也十分容易。正是由于 HTML 的简单性和易用性，使 Web 技术及其应用得到了空前的发展，并很快从计算机界走向全社会，并最终走向千家万户。

HTML 使用各种定义好的标记来分别标识和格式化不同的网页元素，使得网页制作与 Web 信息发布变得非常简单。现在，人们无需了解 HTML 的语法就可以使用各种所见即所得的 HTML 编辑器 (如 Dreamweaver、FrontPage 等) 进行 Web 页面的创作。HTML 已经成为最主要的 Web 页面语言。

下面是一段简单的 HTML 网页代码。

```
<TABLE class=a_1 cellSpacing=0 cellPadding=0 border=1 >
<TBODY>
<TR>
<TD><A href="#"><FONT color=blue>LADY GAGA</FONT></A></TD>
<TD><A href="#">欢子</A></TD>
<TD><A href="#">小沈阳</A></TD>
<TD><A href="#">郑源</A></TD>
</TR>
<TR>
<TD><A href="#">韩宝仪</A></TD>
<TD><A href="#">王菲</A></TD>
<TD><A href="#">小虎队</A></TD>
```

```

<TD><A href="#">姜玉阳</A></TD>
</TR>
<TR>
<TD><A href="#">伍佰</A></TD>
<TD><A href="#">刘若英</A></TD>
<TD><A href="#">梁静茹</A></TD>
<TD><A href="#">谢军</A></TD>
</TR>
<TR>
1<TD><A href="#">张靓颖</A></TD>
<TD><A href="#">王心凌</A></TD>
<TD><A href="#">阿宝</A></TD>
<TD><A href="#"><FONT color=red>更多推荐歌手</FONT></A>
</TD>
</TR>
</TBODY>
</TABLE>

```

上面的 HTML 网页代码在页面中显示了一个 4 行 4 列的表格，表格中有歌手的列表。关于 HTML 中代码的含义，本章后面的内容中有具体的介绍。

然而，以目前的发展来看，HTML 存在着相当大的局限性。由于标准的 HTML 标记已经由 W3C 预先确定，不能根据需要自行定义，所以当描述具有各种复杂内容的文档时，HTML 就显得力不从心。另外，HTML 面向的是数据格式的描述，而非面向数据对象本身，因此，HTML 标记并没有给出数据内容的含义。

使用目前的 HTML，网页文档开发者必须要对文档进行许多调整才能兼容各种流行的浏览器。除此之外，由于浏览器不去检查网页中错误的 HTML 代码，因而导致 Internet 上有大量的文档包含了错误的 HTML 语法。这个问题越来越严重，于是 W3C 开始寻找各种解决方案，XML 就是最好的选择。

1.1.3 可扩展标记语言（XML）

XML（eXtensible Markup Language），就是可扩展标记语言，它与 HTML 一样，都出自于 SGML 标准通用标记语言，但 XML 是 Internet 环境中跨平台的、依赖于内容的技术，是当前处理结构化文档信息的有力工具。在 W3C 组织领导下的小组开发并支持 XML 技术，使用它来标记和处理通过互联网传输的文档信息。

1998 年，W3C 发布了 XML 1.0 规范。与 SGML 一样，XML 让人们在描述特殊文档时可以创造自己的元素标记。此外，XML 的语法还提供了比 SGML 少得多的选项，从而使人们可以更容易地创建和阅读 XML 文档，并且让程序设计人员更容易编写出用来处理和显示 XML 文档信息的应用程序和浏览器软件。

XML 是一种年轻的元语言，或者说是一种定义标记语言的语言，人们在其基础上开发了适合各个领域内部文档发布和数据交换的 XML 应用标准，如数学标记语言（MathML）、化学标记语言（CML）、无线标记语言（WML）等。

XML 不仅可以用来描述网络上的新型文档格式，而且适用于描述任何结构化或者半结构化的数据。结构化的数据包括数据库中的数据、电子表格数据、程序配置数据等；半结构化的

数据则包括 Web 页面、商务文档和各种领域的专业文档等。

XML 满足了 Web 内容发布与交换的需要,适合作为各种数据存储与共享的通用平台。Oracle、IBM 以及 Microsoft 公司都积极地投入人力和财力,研发与 XML 相关的软件,这无疑确定了 XML 在 IT 业界的美好前景。例如,Microsoft 公司的 Microsoft Office、Microsoft SQL Server 和 Microsoft.NET Framework 都允许开发者和终端用户使用 XML 作为网络信息和其他数据的保存与传输格式。

XML 和 HTML 都属于 SGML 的子集,但与 HTML 不同,XML 有 DTD,可以像 SGML 那样作为元语言来定义其他文件系统,而 HTML 和由 XML 派生的 XHTML 都是实例符号化语言,不能去定义其他文件系统。

以下就是一个典型的在新闻发布系统中保存新闻信息的 XML 文件。

```
<?xml version="1.0" encoding="gb2312"?>
```

```
<News>
```

```
<科学技术>
```

```
<ID Name="13">
```

```
<Num>13</Num>
```

```
<Title>英特尔 CTO 访华</Title>
```

```
<Content>
```

金融危机以来,全球经济何时走出低谷成为各方关注的焦点。英特尔刚刚公布的第三季度财报显示,其收入达到 94 亿美元,比上一季度增长 17%,同时也创下了 30 多年来第三季度同第二季度相比之最高环比增幅,这一超出此前分析师预期的业绩被认为是周三全球股市上涨的最大动力。

```
</Content>
```

```
<Type>国际新闻</Type>
```

```
<IssueDate>2009-10-17 13:34:02</IssueDate>
```

```
</ID>
```

```
<ID Name="14">
```

```
<Num>14</Num>
```

```
<Title>联通版 iPhone 月底开售</Title>
```

```
<Content>
```

从与中国移动开始接洽,到谈判出现裂痕,再到中国联通的介入,苹果 iPhone(手机上网)手机在中国大陆市场的推出可谓一波三折,屈指算来足有两年的时间。如今,中国联通终于正式宣布 10 月 30 日在中国大陆开始发售,苹果 iPhone 的神秘面纱也将被揭开。

```
</Content>
```

```
<Type>国际新闻</Type>
```

```
<IssueDate>2009-10-17 13:34:02</IssueDate>
```

```
</ID>
```

```
</科学技术>
```

```
<世界体育>
```

```
<ID Name="25">
```

```
<Num>25</Num>
```

```
<Title>小皇帝病愈恢复训练</Title>
```

```
<Content>
```

詹姆斯确诊患上甲流后一直在家中休息,与其同命相连的还有他的队友达内尔·杰克逊和考比·卡尔,但值得庆幸的是,这次意外并没有殃及到整支球队。昨日,詹姆斯也恢复了训练。“现在发生还算好的,我不希望任何人生病。”詹姆斯表示,他祝愿所有人都健康。

```
</Content>
```

```
<Type>国际新闻</Type>
```

```

<IssueDate>2009-10-17 13:34:02</IssueDate>
</ID>
<ID Name="26">
  <Num>26</Num>
  <Title>马家军死于兴奋剂? </Title>
  <Content>
    1998年,赵瑜的《马家军调查》在《中国作家》整本独卷发表,该书披露了当年红极一时的
    马家军诸多不为外界所知的秘闻,人们看到一个真实的、褪去光环的马家军,这本书也成为中国
    体育报道史上最具轰动性的作品之一。
  </Content>
  <Type>国际新闻</Type>
  <IssueDate>2009-10-17 13:34:02</IssueDate>
</ID>
</世界体育>

```

上面的 XML 文件中保存了科学技术和世界体育两个新闻栏目中的各两条新闻的基本信息。对于 XML 文件的具体使用,在后面章节中会有详尽的介绍。

1.2 HTML 概述

HTML 能构成网站的页面,是一种表示 Web 页面符号的标记性语言。设计 HTML 语言的目的是为了能把存放在一台电脑中的文本或图形与另一台电脑中的文本或图形方便地联系在一起,形成有机的整体,人们不用考虑具体信息是在当前电脑上还是在网络的其他电脑上。我们只需使用鼠标在某一文档中单击一个图标,Internet 就会马上转到与此图标相关的内容上去,而这些信息可能存放在网络的另一台电脑中。

标记是 HTML 语言中最基本的单位。HTML 的标记分单标记和成对标记两种。成对标记是由首标记<标记名>和尾标记</标记名>组成的,成对标记只作用于这对标记中的文档。单独标记的格式为<标记名>,单独标记在相应的位置插入元素就可以了,大多数标记都有自己的一些属性,属性要写在始标记内,用于进一步改变显示的效果,各属性之间无先后次序,属性是可选的,也可以省略而采用默认值。当浏览器收到 HTML 文件后,就会解释里面的标记符,然后把标记符相对应的功能表达出来。

1.2.1 HTML 基本结构

HTML 文本由 html、head、title 和 body 4 个标记组成,这 4 个标记能构成整个网页的效果。它们属于基本标记,其基本格式如下:

```

<html>
  <head>
    <title></title>
  </head>
  <body></body>
</html>

```

实例 1-1 构建网页最基本的结构

在这个实例中，我们通过最简单的编辑工具，来创建一个简单的 html 页面来显示一些文本，具体步骤如下。

01 打开 Windows 中的记事本，创建一个 html 文件，将其命名为 web.html 并在其中输入如下的代码：

```
1.  <html>
2.    <head>
3.      <title>我的第一个网页</title>
4.    </head>
5.    <body>欢迎光临我的网页！</body>
6.  </html>
```

上面的第 1 行与第 6 行代码中的 `<html>` 和 `</html>` 放在文档的最外层，文档中的所有文本和 html 标记都包含在其中，它表示该文档是以超文本标识语言（HTML）编写的。

第 2 行与第 4 行的 `<head>` 和 `</head>` 是 HTML 文档的头部标记，在浏览器窗口中，头部信息是不被显示在正文中的，在此标记中可以插入其他标记，用以说明文件的标题和整个文件的一些公共属性。

第 3 行，`<title>` 和 `</title>` 是嵌套在 `<head>` 头部标记中的，标记之间的文本是文档标题，它被显示在浏览器窗口的标题栏中。

第 5 行，`<body>` 和 `</body>` 标记一般不省略，标记之间的文本是正文部分，一个网页所要显示的内容都是放在这里的，这里也就是我们所要设计的主体部分，也是 HTML 页的一个重要部分。

上面的这几对标记在文档中都是唯一的，head 标记和 body 标记是嵌套在 html 标记中的。

02 使用 IE 浏览器运行该文件，可以看到如图 1-1 所示的运行结果。

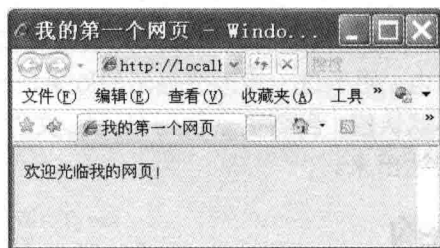


图 1-1 HTML 页面效果

1.2.2 HTML 常用标记

要学好 XML，首先要学习 HTML，熟悉 HTML 的各种常用标记。在表 1-1 中，我们列出了 HTML 常用格式标记及其用法。